

METHOD AND SYSTEM OF LANGUAGE DETECTION

Shamim A. Alpha

ABSTRACT

A system and method for determining the language of an unknown document is provided. For a set of candidate languages, a negative assumption is set for each candidate language that the document is not that language and the system attempts to prove the negative assumption is wrong. If the negative assumption fails for one language, then the document is identified as being in that language. The present system and method provides a higher degree of accuracy when determining the language of a document.